DOCUMENT RESUME

ED 264 298 TM 860 002

TITLE	"Write on Illinois!" A User's Guide to Scoring Student Essays.
INSTITUTION	Illinois State Board of Education, Springfield. Evaluation and Assessment Section.
PUB DATE	Dec 84
NOTE	65p.
PUB TYPE	Guides - Non-Classroom Use (055) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Educational Assessment; Educational Methods; Elementary Secondary Education; *Essay Tests; Evaluation Criteria; *Evaluation Methods; Rating Scales; Scoring; *Scoring Formulas; State Programs; *Student Evaluation; *Writing Evaluation
IDENTIFIERS	*Analytical Scoring; *Illinois Inventory of Educational Progress

ABSTRACT

This report identifies the methods and issues in the evaluation of student writing in classrooms and large scale assessments using an analytical scoring system from the Illinois Inventory of Educational Progress (IIEP). This guide is written for staff in state offices of education conducting large-scale writing assessments, and also for personnel in colleges of education and local school districts interested in implementing a student writing program. This guide includes the following: (1) purposes of assessing writing; (2) the rating guide for functional writing (discourse aims assessed, writing prompts, and rating criteria); (3) specific elements of the IIEP rating criteria and scoring guide; (4) six-point rating scale; (5) training system; (6) classroom adoptions and options; (7) recruitment and selection of scorers for the IIEP; (8) observations and results of the training and scoring process; (9) student outcomes of the writing assessment; (10) assessment outcomes related to instruction; (11) remarks and directions on future research; and (12) bibliography. Appendices which make up half of the document include: (1) Rating Guide for functional writing and functional writing/narrative; (2) writing prompt; (3) training paper example along with feedback; (4) examples of score points from student essays; (5) supplementary scoring materials; (6) score sheets; (7) check paper and feedback; (8) logistics schedule; (9) example of a well developed paper; (10) glossary for the writing assessment; and (11) an outline guide for the writing and assessment. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view. CS

Write on Illinois!

A user's guide to scoring student essays

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Gill, D.G.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 860 002

"Write On Illinois!"

A Users Guide To Scoring Student Essays

Illinois State Board of Education

Program Evaluation and
Assessment Section

Springfield, Illinois

December, 1984

FOREWORD

This Users Guide was developed as a result of the 1983 Illinois Inventory of Educational Progress Writing Assessment. The State Board of Education endorsed a set of recommendations including the writing and dissemination of this guide to assist local school districts with their individual writing programs. This publication is part of a statewide technical assistance plan.

Dr. Edys S. Quellmalz, University of California, Center for the Study of Evaluation and Stanford University is the major consultant and author of the Illinois Functional Writing Scale and portions of this report. For additional information contact Carmen Woods Chapman, Coordinator and Director of the Illinois Writing Assessment. (217/782-4823)

Ted Sanders
State Superintendent of Education

TABLE OF CONTENTS

Introduction. 1

Purposes of Assessing Writing 2

The Rating Guide for Functional Writing 3

Discourse Aims Assessed 3

Writing Prompts 3

Rating Criteria 4

Specific Elements of the IIEP Rating. 5
Criteria and Scoring Guide

The Six-Point Scale 6

Training System 9

Classroom Adoptions and Options 10

Recruitment and Selection of Scores for the IIEP. 10

Observations and Results of the Training. 11
and Scoring Process

Conclusions 15

Student Outcomes of the Writing Assessment. 16

Assessment Outcomes Related to Instruction. 18

Some Random Remarks and Directions on 21
Future Research

Bibliography. 23

Appendices. 25

Glossary for Writing Assessment 53

Outline Guide for Writing Assessment. 55

"Write On Illinois!"
A Users Guide to Scoring Student Essays

The purpose of this report is to identify the methods and issues in the evaluation of student writing in classrooms and large scale assessments using an analytical scoring system from the Illinois Inventory of Educational Progress. The information contained here is based on extensive research and the experience of the Evaluation and Assessment Section of the Illinois State Board of Education.

This guide is written not only for staff in other state offices of education conducting large-scale writing assessments, but also for personnel in colleges of education and local school districts who are interested in implementing a student writing program. A complete set of materials used for the Illinois Writing Assessment is included in the appendices.

Introduction

The Illinois Writing Assessment Program is philosophically committed to linking classroom instruction and the assessment process. The Illinois program, which assesses students at grades four, eight, and eleven, provides flexibility in both the assessment process and use of the student results. The statewide sampling procedure allows the state program to commit resources to instructional research activities not always viable for states with competency testing requirements.

Illinois used an analytical scoring system to assess student writing and is currently involved in follow-up studies with other states on the impact of the training and scoring on individual scorer's teaching practices. Some states have used essentially the same writing assessment with minor differences in the scoring scale which may provide an additional basis for future projects. The emphasis on cooperation and joint research projects rather than merely comparing student achievement data may ultimately contribute to improving the assessment process.

The Illinois assessment took place in the spring of 1983 with the scoring completed by that fall. Most of the data have been analyzed and reported to the State Board and regional and statewide organizations. Future plans call for a repetition of the same prompt for a writing assessment in the spring of 1985, as well as the field-testing of ten additional assignments in persuasive, explanatory, narrative and specific content writing.

The major focus of any large-scale assessment needs to be the improvement of classroom instruction for students. In order to be successful the assessment should meet the following criteria: (Quellmalz, 1984)

1. The assessment clearly defines what it is and is not testing.

2. The assessment provides a credible representation of the defined writing skill status and growth.

3. The methods used to construct, conduct, and verify the assessment meet technical standards.

4. The assessment provides information that will be used at the classroom and system levels to improve writing skill.

The Illinois Inventory of Educational Progress (IIEP) first included direct writing assessment as part of its annual testing program in 1983. The IIEP is designed to assess the knowledge and skills of students on curricular bench marks, to chart trends across years, and to identify factors that may affect student achievement. Each year the program randomly selects approximately 7,200 fourth-, eighth-, and eleventh-grade students from 120 schools to participate in the testing. The IIEP is neither mandated by statute nor connected to state or local high school graduation requirements.

The Rating Guide for Functional Writing (Appendix A.1) that was designed to meet the needs of the IIEP drew upon the theoretical and empirical work of the Writing Assessment Project at the Center for the Study of Evaluation at University of Cal., LA. Funded in part by the National Institute of Education, the project developed over ten studies from 1978-1983 which examined ways to improve the design of writing assessment systems.

This guide includes information on how to select a writing assignment (prompt) and the criteria for scoring. Information will be presented on using the rating guide in the classroom, training and scoring large-scale assessments, and determining instructional implications.

Purposes of Assessing Writing

It is important for any assessment procedure to include a definition of what is and what is not being tested. This definition should reflect the specific component(s) of the writing domain, the mandated (or implied) purposes and the logical and economic constraints (Quellmalz, 1984).

The purpose of the assessment was to describe the current status of Illinois students' writing skills and to begin the process of monitoring skill development over time. To accomplish these goals, policy makers and curriculum experts sought a scoring system that would provide not just a global evaluation of students' composition, but a descriptive profile of students' command of fundamental techniques of clear writing. The detail of the developmental profile was needed to help policy makers allocate resources and understand the rationale for the allocation and to help program personnel determine areas of instructional need.

The Rating Guide for Functional Writing

The Rating Guide for Functional Writing (Appendices A.1 & A.2) combines two approaches to evaluating student compositions--holistic and analytic judgments (Quellmalz, 1983). The guide provides analytic ratings for four elements of clear writing: focus, organization, support/elaboration and mechanics. The mechanics category is further divided into sentence construction, usage, spelling, punctuation/capitalization and format. The guide also provides a focused, summary judgment score of how well the composition orchestrates each of these four fundamental elements to meet the rhetorical requirements of the assignment. Criteria for the guide are deliberately limited to the evaluation of a particular range of discourse structures and a specific set of elements within those forms of composition.

Discourse Aims Assessed

Within the last decade, a number of classification mechanisms have been proposed to replace the traditional types of writing: description, exposition, argument, and narration. Two of the most influential schemes come from James Kinneavy's (1971) theoretical work on the aims of discourse and from James Britton's (1976) extensive analyses of the kinds of writing required of students in school. Kinneavy's discourse aims include expressive, literary, persuasive and referential writing. Britton's categories are expressive, poetic and referential tasks where writers explain points of view, events or ideas in order to inform or convince the reader. These kinds of tasks tend to be functionally or practically oriented. They include the kinds of writing students and adults often use in their school work, jobs and personal business transactions; thus, development of functional writing skills is one of the important goals of schooling.

Writing Prompts

One of the most critical tasks of the writing assessment process is the development and/or the selection of the writing assignment. There are a number of elements which can result in variable student writing performance such as the discourse aim, its implied method of development (persuasive, expository, narrative), the topic, the writer's role, and the audience.

Students also need to have sufficient background about a topic. Sufficient warnings about biased topics and the need for interesting writing assignments appear frequently in the literature. Attention also needs to be given to selecting "ageless" topics which appeal to various maturity levels. In order to compare achievement across grade levels (as in the IIEP), cross-age prompts are given to avoid unintended norming of standards within grade levels.

Assignments can also be improved by including necessary topic information as part of the actual prompt. These may include pictures and selection from a number of topics. Writing assignments may also include reminders to the student about checking whether the composition meets the criteria to be used in the scoring (e.g., "Be sure your composition has a main idea."). These instructions may be in the form of a revision check list.

In addition to specific instruction, the amount of time given for the writing assignment is an important variable. The amount of time may influence whether the student is able to plan, outline, revise, etc. Most large-scale assessments, such as the IIEP, are constrained by certain time limits of the testing process, while teachers have much more flexibility with classroom assignments.

Another prompt element is the format of the student's composition. Format issues include specified length of the essay or special forms such as letters. Some assignments could ask students to provide missing parts to partial formats such as focusing or support statement for a given topic.

Since the IIEP was limited to the collection of a single writing sample in its first direct assessment, the decision was made to use a writing assignment or prompt that would elicit functional writing. (Appendix B) Consequently, the term "functional writing" was used to indicate the kinds of writing assignments the guide was designed to evaluate. Students were asked whether they agreed or disagreed that the television set should be unplugged because of its possible adverse effect on their education, health, and independence.

Rating Criteria

The most important question to answer in the area of rating criteria is not whether one should use holistic, impressionistic or analytic scoring, but rather are the definitions of criteria and the rules for the application of criteria to papers sufficiently clear that teachers, students, and parents can understand how the features of a paper fit within a score category? Equally important, can the rating scale measure individual skill growth?

There are a number of ways to set up the rating criteria for scoring student papers. Existing rating guides differ in how they approach the assessment. One way of looking at a rating guide is to check the parts of writing which have been selected. These include such things as support, organization, and focus. Rating also includes the breadth of the scale or whether the scale measures pass/fail or allows for a full range of scores from poor to excellent. The precision of the criteria definitions are also important. Scale criteria may be general or specific and apply to all types of writing or only to a particular prompt. The type of score format (e.g., single score, component scores or a combination of scores) is another rating decision.

Rating criteria are also concerned with how different age or grade levels are treated in the scoring process. Scoring scales which are limited to a single grade or age level can not be used for tracking skill development or program effects. If the same scale is used at multiple grade levels but the papers are scored separately, the scale may inadvertently become normed within grade levels.

Rating criteria should be used which provide useful information to both teachers and students about their writing strengths and weaknesses. Mixed scoring methods that provide both a summary of the effectiveness of the

total essay and ratings of component skill levels will be useful to teachers.

The scoring of multiple facets of writing has been criticized for the necessity of increased time and cost. Recent reports from large-scale assessments suggest that the scoring time for analytic rating, reflecting good training preparation, is moving much closer to the holistic scoring time. (3 1/2 minutes per paper for the IIEP).

The scoring method used in both large state assessments and local districts sends a message to the classroom about the importance of various aspects of writing. Scoring Guides should reflect policy decisions about the most important aspects of writing at both the state and individual classroom level. Scoring guides with two criteria for text-level features (e.g. the essay is clear and well organized) and fifteen criteria for sentence-level conventions imply that mechanics are more important that the organization and elaboration of ideas. Rating guides need to reflect the results of continued research in defining writing components. The need for competency tests to use comparable or uniform criteria from year to year may result in maintaining a problematic scale, which does not result in good information about student writing skills. Joint projects and sharing of information will facilitate continuing improvement in both rating criteria and the training process.

Specific Elements of the IIEP Rating Criteria & Scoring Guide

The guide (Appendices A.1 & A.2) has a number of features that distinguish it from other rating scales. First, the guide is criterion-referenced, i.e., the criteria represent increasing levels of writing skill development. Score categories describe features of compositions that distinguish incomplete or partially formed compositions from those that are more fully developed. Incomplete or partially formed compositions are characteristic of less mature writing. Fully developed compositions reflect the skill of more proficient, mature writers. The guide references describe standards of clear writing regardless of the age level of the writers. Thus, the decision is not that a topic statement is a clear main idea "for a fourth grade paper," rather that it is a clearly stated main point.

A second feature of the guide is that the composition elements included were distilled from analyses of composition features referenced by expert readers and placed on rating scales as important dimensions for evaluating essays (Diederich, et al., 1961; Freedman, 1979; Smith, 1978; Winters, 1978). Third, the guide elements were limited to aspects of compositions that appear frequently in resources for teaching composition and as essay elements that recent composition research indicates can be taught (e.g., Scardamalia & Bereiter, in press).

The four basic elements, then, are:

1) Focus--the clarity with which a composition presents a clear main idea or point of view;

2) Support/elaboration--the degree to which the main point is elaborated and explained by specific detail and reasons;

3) Organization--the clarity of the logical flow of ideas and the explicitness of the text structure or plan; and

4) Mechanics--use of conventions of standard written English.

The guide also calls for a fifth overall scale that represents a global judgment of how effectively the composition as a whole addresses the essay assignment. These features of composition are ones that:

1) appear to be fundamental;

2) differentiate competent from unclear writing; and

3) are amenable to instruction.

The guide does not attempt to specify criteria for other, less reliably defined measured or taught, aspects such as originality, depth of thought, voice or style.

The Six-Point Scale. Each of the elements is rated on a six-point scale whose numbers indicate the level of the composition's development and, by inference, the level of the student's skill development. A score of 1-3 indicates that the element is absent or only partially formed; a score of 4-6 signals that the element has been minimally, well, or more fully developed. A rating of "4" means that the feature appears in a "bare bones," basic, but acceptable form.

Each element is rated independently, with the exception of the Overall score, which is a focused holistic judgment. Mechanics scores of 3 or below receive additional diagnostic checks to identify specific areas of weakness.

One way of understanding the rating scale and the student results of the writing assessment is to look at excerpts from student papers. The focus of the analysis is on whether a student can meet the objectives of clear functional writing in responding to the "television" prompt. (Appendix B)

Students were rated on their ability to express a point of view and maintain that point of view throughout the essay. These examples show at least minimal development of focus. About one-half of the fourth graders were able to meet the criteria for a minimally developed paper by taking a position and maintaining it throughout the paper as shown below.

> "I agree because I think people should get involved in other activities. I think it is wrong to get the wrong ideas about life from television...They use foul language...There are many reasons why children shouldn't watch so much television."

A high school student wrote:

> "I really don't disagree or agree with unplugging all the TV in America...Television could help young people by...I don't think they ought to watch it all the time..."

Another acceptably "focused" paper from a fourth grader:

> "I agree TV should be unplugged. The children couldn't watch it all the time..."

An example of an eleventh grade paper which does not maintain focus is as follows:

> "Because my homework and cheerleading practice, I don't have a lot of time to watch TV...For me television is relaxing...There are many educational shows..."

A fourth grade student paper was rated '1' because neither the subject nor the main point was clear.

> "I do the same thing at my TV but sometimes I cannot look at TV because my mom and dad won't let me and sometimes they leave I turn on the TV and look at the lone ranger and..."

Some fourth grade students met the objectives of organizational skills by having a clear plan and signaling some points with transitions even though in many cases these were with the use of "and."

> "I disagree because I love cartoons and all the funny movies and all the specials...and my dad and I like football games...and that's why I disagree.

A more developed eleventh grade paper which rated a "6" in organization uses a number sequence to set off transitions.

> "I agree that unplugging all the TV sets would improve our health, make us more independent and better our education...
>
> First of all, TV affects our health by... If we watch too much, then... Also ...
>
> Second, we would become more independent if... because... We would be exercising.... so...
>
> Third, we would learn more because...
>
> Therefore, we should unplug..."

The supporting evidence used by the student was rated by the degree to which the main point was elaborated. The following fourth grade paper is a good example of a composition which does demonstrate the ability to meet the

criteria of a minimally developed paper in organization although the student is far from even a minimally developed level in mechanics.

> "I agree T.V. should be unpluged. The children couldn't watch it all the time. If TV. is unpluged they woudn't have anything to set there eyes on for every hour to watch actors preform then theyed do there homework instead of watching T.V. theyed play and get strong thed start doing exercises when they get board thed go to bed at a decent hour at night theyed do there schoolwork even better they wouldnt think abought what they would watch when they get home. Also theyed eat more because theyed work up an appitite. That makes them tree in one person healtyer, stronger better in getting educated and thats why."
>
> The end

Another example of a minimally developed paper in organization is taken from a fourth grade paper with a better command of the use of standard English.

> "I disagree that we should get rid of TV. There are lots of good shows like 1-2-3 Contact, Sesame Street, and Mr. Rogers that teach us things. Fathers would not be able to watch the news or sports and housewifes would be mad because they wouldn't be able to watch their soap operas. TV is great for a family to get together and watch a good movie."

An eighth grade student used much more general support and failed to give second-order elaboration with specific examples to meet the "developing" criteria.

> "I disagree. TV helps people find out what's going on. It shows the weather and also has special educational programs. TV is watched by millions of people each day. I think it would be hard to take all the sets away. It would make kids grouchy and they wouldn't have much else to do but sit around with their homework and twiddling their thumbs."

Another example of a paper which fails to meet "developed support" criteria is shown here. Note that the student begins with saying that, "TV hurts your eyes," but then does not stay with the task of telling the reader why. The student proceeds to tell you what he/she does as opposed to giving reasons for agreement or disagreement with the prompt statement.

> "Watching TV hurts your and my eyes. You can not read very well. So that means your grades drop. And watching TV is bad for your health. I watch TV with my friends. Watching TV is some of my friends hobbies BUT NOT MINE! Watching TV is not really what I have planned when I get home from school. Playing outside is what I have planned when I get home from school. Not watching TV but something I do. But not all the time..."

(See Appendix D for more student examples and Appendix I for a "6," a well-developed paper.)

The Training System

Overview. The training system begins by introducing raters to the scoring guide, providing examples of papers that fit within each score category, and having raters practice scoring papers. All training papers (Appendix C) have been prescored by a validation committee. The training papers represent the pool of papers that will be rated independently and illustrate clear and marginal examples of each score category. At the end of training, raters score a set of pilot papers (qualifying papers) that have also been prescored.

During the independent scoring, raters periodically score a set of "check papers" (validation papers) to be sure that they are continuing to apply criteria accurately (Appendix G). After scoring a check packet, the rater receives a feedback sheet which presents the scores given to the check papers by the validation committee and the rationale for each score (Appendix G). If the rater's score differs from the scores on the feedback sheet, the rater works with one of the training staff before proceeding to rate more papers independently.

The training system used to establish and maintain rating consistency and accuracy differs from some of the methods used in large-scale writing assessment. The most important feature of the training system is its attempt to teach the raters the relationship between the description of each score point and the specific components of the composition that fulfill the scale criteria. This continual pursuit of precision has involved refining language used in the guide to provide descriptions of how a feature would appear at each score point, rather than relying on examples of papers alone to clarify ambiguous rating scale language. Therefore, a "4" in organization is a paper where the plan is clear, many major points are signaled by transitions and/or paragraphs, points are logical and there are few or no digressions. These characteristics of a "4" paper have been laboriously honed through five generations of rating guides and their use in training hundreds of raters to score over 25,000 papers. The precise language for each score category contrasts with scales that describe a paper as "well organized."

Another distinctive feature of the training method is that it treats the goal of establishing and maintaining accurate and consistent use of the guide's criteria as an instructional problem. The task is to teach raters to use criteria within each score category as rules or attributes for classifying elements of compositions as instances of that category. The guide provides explicit rules for classifying elements of composition; the guide does not rely on raters to induce classification rules from examples or anchor papers.

Consequently, the design of the training also differs from many procedures used in large-scale assessment. Following an introduction to the guide's elements and specific score criteria, excerpts from student essays are used to illustrate the use of criteria to classify an element into a score category. This procedure is used separately for each element. Then the trainers use an overhead projector to display papers and walk through the

reading and rating process. Colored marking pens are used to identify the places in the composition that represent the student's attempt to establish focus, provide support, organize and use conventions. This modeling strategy demonstrates the analysis, feature identification and score classification processes and decisions. All discussion focuses on relating aspects of compositions to the language of the scoring guide.

A final, notable characteristic of the training is its stress on the implications of each classification and score point for instruction. For example, an essay that opens by agreeing with a position but ends, in fact, offering reasons for agreement and disagreement would receive a "3." The explanation of the score would be that a clear focus was not maintained, and that the instructional implications of that score would be for the teacher to work with the student on how to construct focusing statements that indeed summarize the reasons given. Such descriptions of a paper's features and how the students could be taught to improve those features help to avoid abstract literary criticism of papers.

Classroom Adoptions and Options

Individual teachers have numerous options available to them in designing their own rating guides. These guides are based on their specific instruction and standards for meeting specified criteria. Information to students on individual strengths and weaknesses of a composition may serve as both teacher-oriented instruction and as self-checks for students as they write and edit their own or peers' papers.

Another use of the criterion-referenced writing assessment is the development of "writing across the curriculum" programs. Teachers from all curricular areas become active in the student writing process as they see specific criteria by which they can measure the writing in their individual subject areas.

Elementary teachers may be particularly interested in the scale and the relevance for teaching students the major components of functional writing. traditional elementary language arts curriculum has been based on the mechanical skills of grammar, vocabulary and spelling at the expense of focus, organization, support or any other content-oriented part of writing. In addition to these instructional implications, teachers find it extremely helpful to have a set of commonly applied criteria for determining the level of student writing.

Recruitment and Selection of Scorers for the IIEP

A major component for the success of the writing assessment involves the recruitment and selection of competent scorers. In the large-scale assessment, there are different options than in the school district where the personnel are fairly well predetermined.

In order to score the 7,200 papers for the IIEP, it was necessary to hire approximately 60 people outside the Illinois State Board staff. Letters were sent to principals of schools which had participated in the IIEP asking

if individuals on their staff would be interested in participating in a "scoring workshop." As a result, a number of teachers filled out preliminary interest forms.

The 7,200 student papers included equal numbers from fourth, eighth and eleventh grades. As noted earlier, papers were scored according to the criteria of the scale rather than normed according to the grade level of the individual student. For these reasons, two-thirds of those selected were secondary teachers with English backgrounds and one-third were elementary teachers. There was also an attempt to achieve some geographical representation reflecting the diversity of Illinois school districts. The recruitment process was hindered by the fact that teachers were inundated with end-of-the-year term papers and tests and not enthusiastic about a summer project which required scoring hundreds of student essays. Approximately two-thirds of the raters were selected from applications, and the remainder were personally recruited from Springfield and surrounding communities.

Ten of the sixty raters were selected to serve as table leaders or assistant trainers and given a day of inservice prior to the regular training. These individuals were responsible for assisting the five to six people who were assigned to their respective tables. Their role was to assist the trainers in responding to questions during the training as well as to help with the logistics of handling 7,200 papers.

Table leaders can provide valuable support during the training of large groups. If the table leader feels comfortable using the rating scale, he/she is able to help the five or six people assigned to the small group during the actual training. These individuals must be well-versed in the use of the rating scale and able to communicate this knowledge to others and generally to assist in the logistics of the training process. Most importantly, these individuals need to be supportive of the philosophy of the scale. Their sense of "ownership" enables them to assist in the difficult first steps of learning the scoring system. The recruitment of these individuals may be easier in a school district where individuals' strengths for this task are better known. In the large-scale assessment at the state level, selection decisions were based more on vita and self-reported biographical information. It may be possible to have a self-selection process based on specific expectations and written job descriptions for the table leaders. It may even be possible to design a pretest to select those who could serve as table leaders. Such a pretest could contain questions on composition, actual editing, and/or an actual essay.

Observations & Results of the Training and Scoring Process

Many methods have been proposed for analyzing writing assessment data in terms of its quality (rating accuracy, consistency, and scale properties) and its meaning (relationship to curricular areas and performance levels) (Quellmalz, 1980a, Quellmalz, 1980b). The check paper procedure was one of the methods used to measure rating quality in the IIEP. These checks permitted trainers to monitor the scoring by comparing each rater's scores

to the criterion scores established by the experts. Results indicate that rater agreement with the twenty-four check papers (within one point) was above the 80% level, with the exception of the Focus scale. The results were: Focus--74%, Organization--86%, Support--81%, Mechanics--85%. Since a good portion of the check papers were "marginal" examples of a score, i.e., the decisions rules were not clear-cut and easy to apply, the levels of agreement are particularly significant (Chapman, Fyans, and Kerins, 1984).

These prescoring procedures seemed to provide a good method to alleviate instructional problems when new raters came to different interpretations on applications of scoring criteria (Quellmalz, 1984). This documentation of decision rules does have some drawbacks as some variations of scoring rules, which may be appropriate, are viewed by some raters as lack of consistency by the trainers.

A second measure of rating quality is inter-rater agreement, i.e., how each rater scored compared with the other person who rated the paper. Percentages of agreement between raters, within one point, were: Focus--76%, Organization--81%, Support--79%, and Mechanics--86% (Chapman, et al.).

A correlational analysis of individual rating consistency by grade level taught, total years of teaching and professional degree (BA, MA, PhD) does not reveal significant differences which can be attributed to specific background variables.

	Overall	Focus	Organization	Support	Mech.
Teaching Experience	.012	-.195	-.087	-.214	-.138
Degree	.233	.025	.156	-.071	.146
Elementary	-.131	.091	.271	.119	.380
Secondary	-.052	-.161	-.075	.035	-.269

The data suggest that regardless of teaching level (elementary or secondary school), years of experience and/or advanced degree, there is no one particular group which demonstrated more consistency with either the expert scored check papers or the other raters. The relative high correlation on the mechanics scores (positive for elementary and negative for secondary) could result from the fact that secondary English teachers probably emphasize and spend more time focusing on students' mechanical errors.

The general guidelines used by the Illinois Assessment staff of securing a group of raters who reflect the percentage of student papers in the sample

(two-thirds secondary and one-third elementary teachers) as well as some geographical considerations appear to be politically sound without sacrificing accuracy or consistency. Additional selection considerations may be taken into account as the assessment process is developed in future years.

A number of individual rater observations, obtained through an evaluation survey following the scoring, were related to the stress and anxiety of learning a new process. A key factor can be traced to the expectations of the individual scorer and how she/he perceived the objectives of the task.

Some of the problems encountered in the Illinois Assessment were the result of recruiting teachers primarily for a workshop to learn a new writing assessment system and secondarily score 7,200 student papers (240 per rater). A workshop format usually promotes participant interaction and solicitation of individual ideas based on professional experiences.

This type of expectation is reflected in the the evaluation survey:

> "...I came for training, not to merely score papers..."
>
> "...As educators, our input was not wanted. In essence, we were to be human computers scoring a number of papers..."
>
> "...We are teachers who are actually teaching these skills. Let us express our reactions. Don't forget that we are professionals -- allow us to function."

Some related concerns were expressed about the lack of understanding of the major objectives of the scoring process itself:

> "I got along a lot better when I finally understood that 'goodness' and 'competency' are not the same thing."
>
> "It should be made clear that you're looking for competency level and not degrees of excellence."

The following is an interesting commentary on the difficulty some teachers experienced in the process:

> "...I least liked sitting still and not talking...remember that most teachers are used to giving directions not listening to them."

Despite these negative comments which were written immediately following the week of training and scoring, over three-fourths of the teachers indicated that not only had they gained valuable information about a new way of evaluating student writing, but they plan to use it in developing writing instruction.

> ...(I) will use a variety of scales with criteria focusing on the objectives of the writing assignments."

"...I have new insight into standards for writing."

"...I was made more aware of student writing problems."

"...(I) will recommend to colleagues to discover how difficult it is to be truly analytical in scoring and yet how practical in diagnosis and in planning instruction the analysis can be."

"I got some much needed help in learning to evaluate student writing."

"I feel that I am more competent in grading or evaluating composition than before I had the workshop."

Another aspect of the training, related to the psychological impact of the process, was the identification and singling out of those individuals who were inconsistently scoring with the check papers. During remediation sessions, trainers used actual student papers which raters had been assigned. The trainer and two or three raters read some of the papers together in order to discern the raters' application of the scale decision rules. Retraining on certain aspects of the scale was then done using the specific examples in the papers. Some of the individuals who were singled out for retraining saw it as a traumatic event. Some expressed humiliation at being so identified in front of their peers.

Remediation is one alternative to eliminating the poorer raters from a large pool of individuals when restricted funding inhibits the hiring of a sufficient number of extra scorers or in an individual school district where the pool of raters is limited to existing staff. If scorers are selected for additional training, information from both check papers and other raters needs to be provided for individuals on their specific scoring problems. Consideration should also be given to the feelings of those who need additional training by offering assistance at the individual level, rather than group remediation.

Despite the fact that scorers were trained to score essays based on a single prompt, it was not possible to meet the quota of 240 papers in 3 1/2 days with 1 1/2 days of training; therefore, another 1/2 day of scoring was added to the schedule. It had been estimated that scoring would take 3 1/2 to 4 minutes per paper. The actual average rate was about 3 1/2 minutes, but the wide range of scoring time among the 60 scorers was not anticipated. Some scorers were taking as much as 7 to 8 minutes per paper, and therefore, the mean scoring time did not accurately reflect the total amount of time needed for all of the raters. Some raters voluntarily graded additional essays which were randomly distributed to them in order for the group to complete as many papers as possible.

An experiment was designed and implemented to test the effect of the feedback sheets from the check papers on "rater drift." All raters were given check papers to score after every ten papers. Two tables of scorers, approximately 12 people, received no feedback sheets after their scoring. Another 12 raters were given feedback sheets on every other check paper they scored (after every 20 papers) and the remainder of the raters received a

feedback sheet following each check paper (after every 10 papers). Preliminary results, when comparing the reliability with the check paper scores, indicate that there is no significant difference between the groups regardless of how often they received the feedback sheets.

Although the results are inconclusive on the effects of the check papers and feedback sheets, there does appear to be some perception on behalf of the scorers that the check papers and feedback sheets were helpful in the scoring process. Over 85% of those who received the feedback sheets found them to be "most useful."

"...The critique sheets kept me on my toes."

"...I did not get the feedback sheets from the check papers so I couldn't check my progress."

"I felt that I needed the check paper feedback to make sure that I was doing OK..."

Conclusions

Many of the analyses of writing assessments have concentrated on the measurement aspects of reliability and validity of the scoring and on student results based on the scoring method. The training of scorers and the ratings of thousands of student essays add new dimensions to the writing assessment. Those raters hired for a large-scale assessment need to understand their specific task and the purpose for which they have been employed. In addition, physical surroundings and pace need to be considered as important components of the process.

The scoring experience appears to have instructional implications for the individual classroom teacher. Most participants in the Illinois assessment indicated that they plan to use a form of the analytic scoring process in their writing instruction. Despite certain discomforts and the tediousness of scoring the essays, teachers saw a practical application for what they had learned.

The Illinois Assessment rater training and scoring process met the predetermined objective of 80% agreement within one point on both expert scored check papers and with another rater on all subscales except for focus (74% and 76%, respectively). Raters were able to use similar criteria for assigning scores across all grade levels. From the Illinois data, it appears that there is no single or combination of group variables which can predict a more accurate rater within the confines of the selection process. The positive results of the training process were reflected in rating quality of both the check papers and the inter-rater agreement. Raters were consistent with each other and to criterion scores over time.

More work on the use of check papers, training to different prompts, time spent on training, and content analysis would all contribute to improving the training process. Refining the process will result in a more accurate and informative assessment process.

Student Outcomes of the Writing Assessment

Table 1.3
Average Performance of Students
on Each Writing Scale

Scales	Grade 4	Grade 8	Grade 11
Focus	3.28	3.73	4.05
Organization	2.83	3.57	3.93
Support	2.72	3.49	3.85
Mechanics	2.72	3.45	3.82

The Focus Skill was rated highest across all three grade levels as seen in Table 1.3. Fourth grade students are better able to focus their arguments than support them or use standard English. Eighth and eleventh grade students had similar profiles across the writing scales.

The aggregate means across all four scales are 2.83 for grade four, 3.50 for grade eight, and 3.85 for grade eleven. The differences among these means indicate growth in performance equal to one full scale point from fourth to eleventh grade. Further analyses indicate that growth is equal to one full standard deviation of development in writing skill, or a 36% positive difference of eleventh grade student performance over fourth grade performance. From fourth to eighth grade, the increase in performance is 23%. The increase from eighth to eleventh grade is approximately 10 percentage points. This indicates that there is not as much improvement in aggregate performance in writing ability from eighth to eleventh grades.

The percentage of ratings of writing samples can be evaluated in terms of four different segments of the scale. Percentages can be determined for students receiving ratings of 0 to 1.5, 2.0 to 3.0, 3.5 to 4.5 and 5.0 to 6.0. These segments represent absent, developing, minimally developed, and well-developed levels of writing achievement. These results are presented in Table 1.4.

Table 1.4

Percentage of Students at Four Levels of Achievement

	Absent	Developing	Minimally Developed	Well-Developed
		Grade Four		
	0-1.5	2.0-3.0	3.5-4.5	5.0-6.0
Focus	6%	41%	49%	4%
Organization	14%	58%	26%	2%
Support	9%	62%	26%	2%
Mechanics	14%	62%	22%	2%
Average	11%	56%	31%	3%
		Grade Eight		
Focus	1%	29%	56%	14%
Organization	2%	39%	48%	10%
Support	1%	39%	48%	11%
Mechanics	4%	43%	42%	12%
Average	2%	38%	49%	12%
		Grade Eleven		
Focus	1%	19%	57%	23%
Organization	1%	25%	54%	19%
Support	1%	27%	51%	22%
Mechanics	3%	30%	46%	22%
Average	1%	25%	52%	21%

*Numbers may not sum to 100% due to rounding

The results in Table 1.4 indicate that approximately 33% of the fourth graders, 60% of the eighth graders and 73% of the eleventh graders have average scores at the minimally developed level or above (3.5-6.0). At all grade levels, the focus scores are the highest and mechanics, the lowest. Fourth grade students scored almost twice as high on focus as in any of the other scales.

Another aspect of this writing assessment data is the estimated relationship among the writing scales themselves. The intercorrelation coefficients among these four scales were all quite high at each grade level. That is, there is a strong positive relationship between and among the scales. The highest correlations were among the skills of Focus, Support, and Organization. These correlations, as well as the factor analysis results, indicate that the writing results can be aggregated across the scales for each student to produce a writing ability score. By aggregating across all

writing scales, one obtains a cumulative indicant of the student's writing potential. Taken together, these scales are quite powerful in describing the student's ability. (No one would even think of measuring math or reading with just one item).

Assessment Outcomes Related to Instruction

The Illinois Writing Assessment has generated several outcomes which may influence the individual classroom level of writing instruction. Initially, the major effect of the assessment itself is a level of awareness about the area of student writing. The analytical analyses and reporting of results to the State Board of Education generated publicity about student writing ability, as defined by a functional writing scale, and interest on the part of policy makers on the instructional implications of the various profiles of student data. It served as a stimulus for discussions on the importance of writing, the specific components of composition, and teacher inservice. The results of the student questionnaire on the amount of writing required for English, other classes, and for what purposes also brought attention to the fact that (1) students are not writing very much, (2) those who watched less television wrote better, and (3) those who wrote more for their own purposes wrote better.

A second level of influence is on the teacher scorers themselves. In an evaluation survey immediately following the training and scoring session, over three-fourths of the Illinois teachers indicated that not only had they gained valuable information about a new way of evaluating writing, but they planned to use it in developing future writing instruction. Elementary teachers appeared to be particularly interested in using the analytical scale to teach the major components of functional writing. They responded that they found it extremely helpful to have a set of commonly applied criteria for determining the level of student writing in each of the four areas of focus, organization, support and mechanics.

Another survey was sent to scorers asking whether the analytic scoring experience had had any effect on their instruction and/or curricular development during the 1983-84 school year. The preliminary results suggest that there been an increase in the frequency of writing by the students of these teachers. Some teachers felt that they were better able to instruct students in the individual components of a writing assignment and show students how to combine these components in longer exercises. Other scorers suggested that most teachers need more focussed instruction in the teaching of writing. The survey data indicate that the training and analytical scoring process had a positive impact on writing instruction.

The third and possibly the most important result of the state-level writing assessment on classroom instruction is what may be termed the networking and dissemination of assessment information to groups most concerned with preservice and inservice of teachers. As a result of State Board staff participation in informational programs across Illinois, certain individuals and groups have been identified who are both interested and directly involved in writing instruction. Most importantly, these groups have a vested interest in the State's role and resulting policy decisions. The

State Board writing report and action recommendations served as a catalyst to bring together various regional and statewide groups who had either not been aware or failed to acknowledge other writing programs.

Individuals from these groups are now involved in a joint effort and consequently, an ownership in the 1985 assessment. Each group or project contributes a unique activity to both individual classroom programs as well as a proposed assessment plan which combines the statewide sample assessment with individual classroom programs and assessments. The projects are briefly summarized below.

Illinois Articulation Project -- This project is a result of a writing skills task force appointed by the Joint Education Committee which is composed of officers of both the Board of Higher Education and the State Board. Conferences were convened to focus on improving writing skills by achieving closer articulation across all grade levels. Survey information has been collected on writing programs at all two- and four-year public institutions. The next step will be to conduct a similar survey at the secondary public school level. Based on these data, the project steering committee sees their task as recommending writing policies to all English teachers who will then be in a better position to affect statewide curriculum decisions.

Illinois Writing Project -- This project has provided writing workshops to Illinois teachers in approximately 72 of the 1,010 Illinois school districts for the past six years. The intensive 30-hour structured inservice program provides an opportunity for teachers to participate in their own writing activities after which they are better able to assist their students. The project intends to incorporate the analytical scoring method from the Illinois Functional Writing Scale as one component of its evaluation system.

Illinois Humanities Grant Projects -- Illinois State University, Bloomington-Normal, is currently involved with two projects funded through the Illinois Humanities Council. One is intended to produce curricular materials that will help writing teachers at high schools, community colleges and universities teach their courses more effectively. The project involves a series of team teaching arrangements to help teachers from different grade levels gain firsthand experience with various levels of writing courses.

The other project is aimed at the establishment of cross-disciplinary writing programs, the reinforcement of the skills students develop in writing courses, the

development of writing tasks tailored to different disciplines, and the coordination of cross-disciplinary writing programs at the secondary and post-secondary levels. This is a collaboration project involving five Illinois community colleges and ten public high schools. Administrators of these projects are planning to use the state analytical scale in their evaluation process.

Writing Project - Southern Illinois University, Carbondale--This project incorporates both a writing across the curriculum theme as well as the use of the microcomputer. Local school districts are encouraged to send multidisciplinary teams to the two-week workshop which emphasizes the need for all teachers to teach writing. In addition, teachers are exposed to the uses of the computer for writing instruction, which may be particularly helpful to the English staff who have had little computer experience. The project directors believe that teachers should have their own theoretical framework for writing instruction in order to help students learn to write.

In addition to these specific projects, colleges and universities have shown increasing interest and have developed activities for both preservice of teacher candidates and more inservice for their faculties and graduate students. The existence of the Illinois Writing Assessment and subsequent reporting on the instructional use of the analytical scale has generated numerous requests for participation by assessment staff in these inservice activities.

Individuals from the various writing groups are now actively involved in joint efforts with the State Board to promote activities to improve classroom writing instruction. They see the state as a clearinghouse and diffusion center which enables them to better promote writing instruction. As a result, the 1985 Illinois assessment proposal calls for a plan which will continue to sample and analytically score student writing in three discourse modes. In addition, classroom sites which have effective instructional programs at the fourth, eighth and eleventh grade levels will be identified by writing project personnel. These programs will be described and evaluated using a number of different instruments including the statewide writing prompts and analytical scoring scale. Not only will this allow for the focus to be on instruction and application of evaluation instruments, but the report will provide information to other teachers who will be able to adopt or adapt strategies enabling them to improve their writing instruction.

The advantages of this decentralized approach to the improvement of student writing are numerous. The focus is on the classroom level. Teaching practices can be modeled based on unique characteristics of instructors and students. Multiple methods of evaluation may be available including the state analytical model based on a sound theoretical and measurement foundation. The State Board is able to utilize good existing inservice delivery systems which emphasize the writing process while influencing the

development of an assessment instrument for the product. The importance of this networking with the writing interest groups should not be underestimated. Unless classroom teachers perceive a need for and a possible positive result of curricular change, none will occur. The effect of the analytic approach to large-scale writing assessment will be measured by the applicability to individual classrooms and the needs of individual teachers. By working with the writing instruction groups and incorporating the statewide analytic assessment, the Illinois State Board will be able to give individual teachers a measure by which they can compare their students and at the same time validate the instrument in terms of model practice programs.

Some Random Remarks and Directions for Future Research

Statewide assessments using a representative sample of students have a number of advantages and disadvantages over state-mandated testing programs for every student. The fact that they are generally less in the limelight allows for the inclusion of some learning outcomes that not every school system is addressing at the time of the assessment. Furthermore, because only a small sample of students is involved, the state can experiment with innovative and often more costly approaches. However, on the flip side, because only a small sample of students is involved, the assessment generates only a picture of the state. It is not possible to report accurately at the classroom, school or district level. Therefore, the schools in which students are tested do not receive any data and there is no real ownership of the statewide results by any district.

It would seem that one way to increase the instructional follow-through in statewide assessments is to develop sampling plans which would at the very least, test intact classrooms and permit school districts to piggyback on the assessment with enough students to adequately represent the achievement of the school and possibly the district. This would certainly increase the potential instructional utility of having participated in the statewide sample as districts would receive test scores.

In any program which reports data to teachers, there are two procedures which might be tried to increase the probability of instructional follow-through. The first of these is to share the scoring criteria with teachers at the time of testing and ask them to predict how well their students will do on the test. These predictions could be returned to the teachers with the test results. The curiosity factor might prove to heighten teachers' interest in looking for the degree of overlap between their predictions and the students' scores. It also would be interesting to investigate the effect of returning students' papers with the test scores. Particularly in situations in which teachers have made predictions, it would be important to enable teachers to look at students' papers when the predicted achievement is discrepant from the reported achievement.

Testing intact classrooms, securing teachers' predictions, and returning students' papers (separately or in concert) may improve the likelihood of instructional follow-through. However, the most important component is

monitoring the nature of writing instruction. It is necessary to find out what kind of instruction and feedback teachers provide in writing. Are students taught what the elements of good writing are? Are they shown what good focus, organization, and support look like? Are the students taught the importance of attending to audience? These macro-level features often take second place to the more micro-level features of mechanics and usage. How often do students write? Do they receive feedback on the strong as well as weak aspects of their writing?

Bibliography

Britton, J. (1978). "The Composing Process and the Functions of Writing." In Cooper and Odell (Eds.), Research and Composing: Points of Departure (Chapter 2). Urbana, IL: National Council of Teachers of English.

Chapman, C. W. (1984). "Factors Relating to Training of Writing Assessment Scorers: Raters of the Lost Art." Unpublished paper, American Education and Research Association, New Orleans, LA.

Chapman, C. W., Fyans, L. J. & Kerins, C. T. (1984). "Writing Assessment in Illinois." Educational Measurement (3, 24-26).

Diederich, P. B., French, J. W. & Carlont, S. (1961). Factors in Judgments of Writing Ability. Princeton, New Jersey: Educational Testing Service.

Freedman, S. (1979). "How Characteristics of Student Essays Influence Teachers' Evaluation." Journal of Educational Psychology.

Kinneavy, J. R. (1971). The Aims of Discourse. Englewood Cliffs, NJ: Prentice Hall.

Pitts, M. (1978). "Relationship of Classroom Instructional Characteristics and Writing the Descriptive/Narrative Mode." Center for the Study of Evaluation, University of California, Los Angeles.

Student Achievement in Illinois, An Analysis of Student Progress. (1984). Illinois State Board of Education, Springfield, IL.

Quellmalz, E. S. (1984a). "A System for Analyzing Writing: An Example of Research in the Service of Policy and Instruction. "American Education and Research Association, New Orleans.

Quellmalz, E. S. (1984b). "Toward Successful Large-Scale Writing Assessment: Where Are We Now? Where Do We Go From Here?" Educational Measurement, 3:29-32.

Quellmalz, E. S. (1983). "Rating Guide for Functional Writing." (Report to the Illinois State Board of Education). Los Angeles, CA: University of California Center for the Study of Evaluation.

Quellmalz, E. S., Capell, F. & Chou, C. P. (1982). "Effects of Discourse and Response Made on the Measurement of Writing Compliance." Journal of Educational Measurement, 19, 4, 241-258.

Quellmalz, E. S. (1981). "Report on Conejo Valley's Fourth-Grade Writing Assessment. "Center for the Study of Evaluation, University of California, Los Angeles.

Quellmalz, E. S. (1980a) "Controlling Rater Drift." Los Angeles, CA: University of California, Center for the Study of Evaluation.

Quellmalz, E. S. (1980b) "Problems in Stabilizing the Judgment Process." Los Angeles, CA: University of California, Center for the Study of Evaluation.

Quellmalz, E. S., Smith, L. S., Winters, L. & Baker, E. (1980). Characterization of Student Writing Competence: An Investigation of Alternative Score Systems. Paper presented at NCME, Boston. (Grant No. OB-NIE-G-79-0213) Center for the Study of Evaluation, University of California, Los Angeles.

Scardamalia, M. and Bereiter, C. (in press). "Writing Instruction." In M.C. Wittrock (Ed.). Handbook of Research on Teaching. 4th Edition.

APPENDICES

A. 1. Rating Guide for Functional Writing
   2. Rating Guide for Functional Writing/Narrative

B. Writing Prompt

C. Training Paper and Feedback Sheet

D. Examples of Score Points from Student Essays

E. Supplementary Scoring Materials

F. Score Sheets

G. Check Paper and Feedback Sheet

H. Logistics Schedule/Checklist

I. Example of a '6' Student Paper

BEST COPY AVAILABLE

7/13/83

RATING GUIDE FOR FUNCTIONAL WRITING

OVERALL	1. Absent	2. Attempted	3. Developing	4. Minimally Developed	5. Developed	6. Well Developed
						* Stands alone
	* Subject unclear	* Subject clear	* Subject clear	* Subject clear	* Subject clear	* Subject clear
FOCUS						
	* No main point	* Main point not clear or more than one	* Main point not clear	* Main point clear	* Main point clear	* Main point clear
					* Descibes number or type of key points actually developed	* All key points identified/ specified actually developed
(Ignore titles completely)			Only partially summarizes keypoints actually developed			
				* May have a related beginning and end	* May have a related beginning and end	* Beginning and end relate
	* No plan	* Plan attempted, difficult to decipher	* Plan noticeable, can infer	* Plan is clear	* Plan is clear	* Plan is clear
			* Uneven use of transitions * May have few or no paragraphs	* Many major points signalled by transitions and/or paragraphs	* Most major points signalled by transitions and/ or paragraphs	* All major points signalled by transitions and/or paragraphs
ORGANIZATION						
	* Almost no logically connected points	* Few logically connected points	* Some logically connected points	* All points logical	* All points logical	* All points logical
			* Some major digressions or excessive elaboration	* Few minor digressions, no major	* Minor digression	* No digressions/ irrelevancies
	* Little or no support or support is confusing support at same level of generality as the main assertion	* Few major points elaborated	* Some major points elaborated	* Many major points elaborated	* Most major points elaborated	* All major points elaborated
SUPPORT	* Redundant/unclear	* May be redundant, may not be precise or clear	* Some elaboration is specific (it may be a list)	* Much elaboration is specific and many points have second order elaboration	* Most elaboration is specific and second order	* All reasons elaboration is specific and second order
	* Many minor errors	* Many minor errors	* Some minor errors	* Few minor errors	* Few minor errors	* One or two minor errors
	* Many major errors	* Many major errors	* Some major errors	* Few major errors	* One major error	* No major errors
MECHANICS/ GRAMMAR	* Very confusing * Difficult to read	* Confusion				

BEST COPY AVAILABLE

MECHANICS KEY

	Sentence Construction	Usage	Spelling	Punctuation/Capitalization	Paragraph Format
Major Errors	* Subject/verb agreement * Run-on * Fragment * Omitted words that interfere	* Homonyms - it/it's, their/there, to/two/too * Incorrect use of common words * Incorrect pronoun usage * Unacceptable slang	* Misspelled common words (counted only once, not homonyms)	* Initial caps (not to include run-ons) * Common proper nouns * Ending punctuation * Apostrophes	* Titles used to delineate paragraph * Paragraphs are numbered * Inconsistency of separation convention (either block or indented)
Minor Errors	* Connectors between clauses * Use of "because", "but", "and", "so" to begin sentences * Omitted words that don't interfere	* Awkward or odd use of words/phrases, but meaning still clear * Vague abstract language * Slang generally accepted verbal vocabulary	* Unusual, less frequent words	* Periods for abbreviations * Commas/in a series, for opening phrases or clauses	(lack of correct paragraph separation is an organization error)

Mechanical errors which could be classified in more than one category should be counted only once (no double jeopardy). e.g. Run-ons count as sentence construction only, not punctuation also. If the same word is repeatedly misspelled or misused, it only counts as one error.

BEST COPY AVAILABLE

APPENDIX A.2
RATING GUIDE FOR FUNCTIONAL WRITING/NARRATIVE

ELEMENT 1: FOCUS

This element examines whether the subject or topic of the composition is clear. It also assesses whether the paper explictly summarizes the main idea or point of view about the subject that is developed in the paper. The most sophisticated paper will announce in its opening lines the subject, main idea or view, and the subtopic that will be developed. The paper will also review key points in a conclusion. Less well developed papers may leave the reader to infer the main point or may develop several distinct points without typing them together in a general statement or set of statements.

6--The essay can stand alone.
The subject is clear.

The main idea or view is stated.
The key points or subtopics that are developed are specifically named.
Opening and closing statements match or logically relate to the text and to each other.

5--The subject is clear.

The main idea or view is stated.
The general number or type of key points or subtopics are mentioned.
Opening and closing statements may relate to follow from each other.

4--The subject is clear.

The main idea or view is stated.
There is no attempt to specify points that are developed.
The beginning and end may relate, but do not contradict each other.

3--The subject is clear.

Opening or closing statements may specify more or fewer points or subtopics than are actually developed in the paper.
The reader must, but can, infer the main idea.

2--The subject is still clear.

There may be more than one main idea developed.
The reader must work very hard to infer a main idea.

1--The subject may be unclear.

There is no discernible main point.

ERIC Clearinghouse for Reading and Communication Skills

ELEMENT 2: SUPPORT

This element focuses on the quality (specificity and amount) of the support elaborating the essay theme. Support provides additional information about a general point. Supporting details are more specific than the generalizations they are intended to develop. Such details elaborate or flesh out and explain the main idea and subtopics through description, example and reasons. The quality of support depends on depth, specificity, accuracy and amount.

Depth of elaboration can be determined by diagramming details provided in descriptions or reasons in an outline or tree format. The more subtopics or branches, the greater the depth of elaboration.

Specificity is a means of judging accuracy and credibility. More concrete details or reasons enable the reader to judge whether statements are factual or widely accepted assertions. Generalizations, assertions and opinions must be backed up.

The amount of support also influences the score. Support scores for longer essays will depend on the proportion of points or subtopics elaborated by more specific detail and reasons. Obviously shorter essays will have fewer opportunities to develop and back up main points.

6--The essay's main idea or view and all major subtopic are elaborated and explained by specific detail.

5--Most major points are elaborated.

Most elaboration is specific and second order.

4--Many major points are further elaborated.

Much of the elaboration is specific.
Much of the elaboration is second order.

3--Only some major points are elaborated.

Only some elaboration is specific. It may be a list.

2-- Support is attempted, but few major points are elaborated.

Little of the elaboration is precise or clear.
The support may be redundant.

1-- There is little or no support.

Support is very confusing or at the same level of generality as the point it is intended to develop.

ELEMENT 3: ORGANIZATION

This element examines whether the composition exhibits a clear structure or plan of development and whether the flow of ideas is logical. The essay's organization has a "vertical" dimension evidenced by the use of paragraphing and transitions to signal the relationship of points to the main idea. The essay's organization also has a "horizontal" dimension evidenced by the connection of one sentence to the next. Fully developed essays will use paragraphs and transitions to signal the plan or text structure. They will also use pronouns, demonstratives, definite articles, conjunctions, repetitions of words, and synonyms to cue horizontal relationships.

Less well developed papers will fail to use paragraphing or will use it inappropriately. These papers may also use few cohesive ties or transitions to cue the logical relationships. Some weak papers may have digressions or the train of thought may resemble stream of consciousness or free associations.

6--The essay plan is very evident.

The plan is signalled by the division of major points into distinct paragraphs.
The plan is also signalled by use of transitions.

5--The plan is clear.

Most major points are separated into paragraphs and signalled by transitions.

All points are logically developed and related to each other.
There may be a few minor digressions, but no major ones.

4--The plan is clear.

Many major points are signalled by transitions and developed in paragraphs.

Most points are logical.
There may be a few minor digressions, but no major ones.

3--The plan is noticeable, but the reader must infer it.

Only some major points are signalled by transitions.

There are some logically connected points.

There may be some major digressions.

2--A plan is attempted, but the reader must work very hard to infer it.

There are few or no transitions signalling major points.

There are few logically developed points.

1--There is no evidence of a plan.

Almost no points are logically related.

ELEMENT 4: GRAMMAR/MECHANICS

Errors in grammar or mechanics should be weighted according to how seriously they interfere with communication. A gross error, such as a run-on sentence or a sentence fragment, makes it difficult to understand the writer's message. A statement such as, "Tomorrow, I went to the store," also confuses the reader by obligating him to decipher which time context -- future or present -- the writer actually intends. By contrast, a minor error does not seriously interfere with the writer's message. For example, in the sentence, "He going to the store now," the message is clear but the grammar is incorrect. An overabundance of minor errors, however, may seriously detract from the reader's understanding. The intent in this element, is to evaluate errors in relation to how much they interfere with the writer's effectiveness in communicating rather than to attempt to assign different values to the myriad of possible grammatical and mechanical errors that can occur.

6-- There are few or no minor errors. There are no gross errors.

5-- There may be a few minor errors, but no more than one gross error.

4-- There are a few common errors. A few may be gross.

3-- There are numerous minor errors and some gross errors. Sentence construction is below mastery.

2-- There are many gross errors, causing some confusion.

1-- Errors are so numerous and serious that they interfere with communication.

BEST COPY AVAILABLE

Examples of Mechanics/Grammar Errors

	GROSS	MINOR
	No Double Jeopardy on Errors	
Sentence Construction	. tenses . run-on . fragments . subject/verb agreement	. connectors between clauses . "of", "but", "and", "so" (to begin sentences)
Usage	. incorrect use of common words . mixing languages . homonyms (its/it's, their/there, to/too/two) . pronoun reference	. awkward usage, but meaning clear . vague, abstract language
Spelling	. common words (not including homonyms) . misspellings that impede communication	. unusual words
Only one error counted even if misspelling repeated.		
Punctuation/ Capitalization	. end punctuation missing (not including within run-ons) . apostrophes in contractions . commas missing so communication impeded . initial capitals (not including run ons) . common proper nouns	. commas in series or after introductory word groups . periods for abbreviations
Paragraph/ Format	. titles . numbers . inconsistent method of separation . no method of separation	. use of spacing instead of indentation (as in business letters)

BEST COPY AVAILABLE

APPENDIX B

1983 WRITING PROMPT

ILLINOIS INVENTORY OF EDUCATIONAL PROGRESS

DIRECTIONS FOR THE WRITING ASSIGNMENT

In the first part of the test you will write on an assigned subject. The subject is described on the next page. You will have about 25 minutes to do the assignment.

You should have enough time to think a little before you write--you may want to make some notes or a list of ideas before you start. As soon as you have your ideas in mind, begin to write. You should write so that someone who does not have a copy of the assignment will understand what you are writing about. Turn the page and read the assignment.

ILLINOIS INVENTORY OF EDUCATIONAL PROGRESS

Writing Assignment

It always strikes me as a terrible shame to see young people spending so much of their time staring at television. If we could unplug all the TV sets in America, our children would grow up to be healthier, better educated, and more independent human beings.

Do you agree or disagree? Explain and illustrate your answer from your own experience, your observations of others, or your reading.

APPENDIX C

Training Paper 1

I disagree strongly. I watch alot of television, and I think that I am very well educated. I think that television has a lot to offer in terms of eduction. If you unplugged all of the T.V. sets in America, the only source of information for students would be textbooks, and radio (which in my opinion is not enough You would be cut off from the recent things happening in the world today. You wouldn't be in touch with the outside world. You would be confined in a sort of way. I think people should watch some television at least everyday. But you should

BEST COPY AVAILABLE

get the better of your education at school.

Some networks have separate educational channels, which I think is a very good idea. They have some very interesting programs.

If you cut off T.V. you wouldn't be able to watch live events. Most of those are educational.

I think the television is an excellant invention

BEST COPY AVAILABLE

APPENDIX C

TRAINING FEEDBACK

Code Number Training Paper 1

ELEMENT	RATING	COMMENTS
FOCUS	6	The position is clear and the main reason, "television has a lot to offer in terms of education" describes the points made.
ORGANIZATION	3	Some points are logically developed, but some points are unclear. There are too many one-sentence paragraphs.
SUPPORT	3	Many reasons are general statements or opinions that need to be backed up. "You would be cut off from..."
GRAMMAR/MECHANICS	3	Two incomplete sentences. Pronouns keep shifting person several usage problems.

Mechanical Skills below mastery

__X__ Sentence Construction

__X__ Usage

_____ Spelling

_____ Punctuation and Capitalization

_____ Paragraph Format

BEST COPY AVAILABLE

APPENDIX D

EXAMPLES OF SCORE POINTS FROM STUDENT ESSAYS

ELEMENT 1: FOCUS

6-- Example #1

Unplugging all the TV sets in America would not make our children grow up healthier, better educated, or more independent human beings...

There are many educational shows on TV such as.....

TV also teaches us how to take care of our bodies.....

TV does not make us less independent.....

So unplugging the TV will not make us better educated, healthier or more independent.

Example #2

I disagree that we should get rid of TV because we can learn a lot from TV.....

Children learn to count by watching.....

Grownups learn about the world by watching the news.....

TV really has many educational programs.....

If you unplug the TV we will not learn as much.

5-- Example #1

I agree that we could be doing many other things besides watching TV all the time.....

We could exercise and .....

If TV were gone, children would go to the library and some would play ball.....

So that's why children shouldn't watch so much television.....

Example #2

I do agree with this theory to some extent, but I also disgree.....

Television is not all bad.....

But people without television are much healthier both mentally and physically.

BEST COPY AVAILABLE

4-- Example #1

I really don't disagree or agree with unplugging all the TV in America.....Television could help young people by..... I don't think they ought to watch it all the time.....

Example #2

I disagree with what is said.....

Example #3

Television is a harmful thing.....

Example #4

I disagree because I enjoy watching TV.....

3-- Example #1

Because my homework and cheerleading practice, I don't have a lot of time to watch TV.....For me television is relaxing..... There are many educational shows.....

Example #2

I agree, but I also like to watch television. And I like to read. There are good television and there are bad ones.....

Example #3

Television makes us healthier and better educated..... Sesame Street teaches..... From the news..... I watched the space shuttle..... So don't unplug the TV.

Example #4

I disagree. Television teaches us many things. There are shows on how to cook.....Sesame Street helps.....I am more independent because ..... and I am healthier too.

2-- Example #1

I agree because.....

I also disagree.....

Example #2

I disagree because some people don't care about their body...and we should not be here now if there weren't no television... children run the street a small amount of people talk to television... they would start to play education movies on your TV set.

1-- I do the same thing at my TV but sometimes I cannot look at TV because my mom and dad won't let me and sometimes they leave I turn on the TV and look at the lone ranger and .....

ELEMENT 2: SUPPORT

Examples of Score Points from Student Essays

6-- Example #1

I don't think that unplugging the TV will make me healthier or better educated.

I learn a lot from television. I watch the educational shows about science on channel 7. I learned how to save a life from watching an emergency show. I learn about the world from the news.

There are exercise shows on TV that help you get strong muscles. There are also shows on how to cook and eat healthy food.

Example #2

I agree with the statement given. I believe kids spend too many hours in front of a television set. Their eyes will probably in years to come, be ruined. Sometimes kids have the set blasting and that doesn't help their ears. Junk food is constantly advertised. It can stunt their growth or give them cancer.

Cartoons are a bad influence on children. Some kids believe it! I know a young boy who used to run into walls because he thought he was Captain Marvel. Also, shows like Bugs Bunny make kids think that hitting and jumping off cliffs is OK.

5-- Example #1

I agree somewhat, but I wouldn't completely get rid of it. We do watch too much television. My nephew is glued to the tube on Saturday mornings and doesn't even want to eat his breakfast. I myself don't watch as much as I used to. I'm too busy.

Television doesn't have all bad programs. Some of them are educational.

I'm not sure we would be healthier. I would rather go outside and play. My sister doesn't watch very much TV now that she's old enough to drive and go out on dates.

4-- Example #1

I disagree that we should get rid of TV. There are lots of good shows like 1-2-3 Contact, Sesame Street, and Mr. Rogers that teach us things. Fathers would not be able to watch the news or sports and housewifes would be made because they wouldn't be able to watch their soap operas. TV is great for a family to get together and watch a good movie.

Example #2

I agree and disagree. There is a lot of sex and violence on TV. Kids get a bunch of ideas, sometimes good, sometimes bad.

Kids spend too much time watching TV at night when they should be doing their homework. I have to finish my homework before the TV goes on.

On TV you can learn a lot about certain problems and how to deal with them. For example, on a show like Different Strokes.....

3-- Example #1

I disagree. TV helps people find out what's going on. It shows the weather and also has special educational programs...TV is watched by millions of people each day. I think it would be hard to take all the sets away. I would make kids grouchy and they wouldn't have much else to do but sit around their homework and twiddling their thumbs.

Example #2

I do agree, but I also disagree. Television is not all bad. There are some very educational programs, although some are definitely not. It is up to the parent and the child to decide which shows to watch. There is usually a channel on a television especially for educational programs.

As for being healthier, I totally agree. People without television are healthier. It is better for a child to find some thing creative to do, or to go outside. Being able to take this initiative on their own builds character and independence.

2-- Example #1

Watching TV hurts your and my eyes. You can not read very well. So that means your grades drop. And watching TV is bad for your health. I watch TV with my friends. Watching TV is some of my friends hobbies BUT NOT MINE! Watching TV is not really what I have planned when I get home from school. Playing outside is what I have planned when I get home from school. Not watching TV but sometimes I do. But not all the time....

Example #2

I agree? I watch TV a lot and eat sweets. I know that it is not healthy but I can't help it. My uncle even said TV is not healthy. But he just still watches it. If somebody would come and unplug our TV I would be mad. I love watching TV. I watch.....

1-- Example #1

I agree because some people stupid and some aren't. Some need to be educated and some don't. If people stop watching TV maybe they can get something done for a change. Some are educated but the rest just want to goof off and watch television. They just watch TV too much. And that's why they don't learn very much. If there were no television no one would be stupid.....

ELEMENT 3: ORGANIZATION

Examples of Score Points from Student Essays

6-- Example #1

I agree that unplugging all the TV sets would improve our health, make us more independent and better our education.....

First of all, TV affects our health by..... If we watch too much, then ..... Also.....

Second, we would become more independent if..... because..... We would be exercising..... so.....

Third, we would learn more because.....

Therefore, we should unplug.....

Example #2

I agree and I disagree that.....TV can be harmful because..... It can cause ..... and ..... However TV can also be beneficial. Shows such as ..... These shows help us to learn..... So there are advantages and disadvantages to unplugging.....

5-- Example #1

TV can help us become better educated and healthier. We would all be much worse off.....

Educational shows like..... are good because..... My favorite one is 1-2-3 Contact. I really like the part where..... I really made me laugh.

Another good part about television is the way it shows us how to stay safe and healthy. Besides that..... So that's why TV shouldn't be unplugged.

4-- Example #1

If you try to unplug everyone's TV, it will be a disaster. For example, young children will not learn their ABC's. And they won't know how to..... Grown up will miss their detective shows and.....

There won't be anything for people to do. Since they are bored, kids will be on the streets, causing trouble. I suppose some kids might go the library, but not any of my friends. We would cause traffic jams because we play ball in the street.

Example #2

I agree with the statement. Turn off the TV and kids will do homework. Turn off the TV and they will get some fresh air. And another thing, they have better eyesight..... So get rid of TV.

3-- Example #1

I agree that TV is bad. Kids are hooked on cartoons. They always..... They should be studying and ..... They don't get any exercise..... The stuff they show on TV is awful too. There is lots of sex and violence..... Like on..... My mom says that these shows are terrible. I'm not allowed to watch them.

Example #2

I think we should leave TV alone. There are lots of good things to watch on TV. You'd have to be crazy to turn it off.

My favorite show is Magnum P.I. My mom likes the soaps and my dad watches the news to get the baseball scores. Old and young people see a different world through television. They see life as fantasy or reality whichever program they choose to watch. Without television some people would not be able to dream wish or express their feelings. But when abused, television can be as dangerous to ourselves and others as a gun, knife or any other form of weapon used to harm ourself or others.

2-- Example #1

Don't turn off my TV. I don't watch much television because I work nights and I am at school all day. I might watch it for a few minutes when I come in from school. If I'm not working I'll usually go hunting with my brother-in-law. On Sundays I go to church in the morning play football in the afternoon and then to to church at night. So I don't watch television very much. I usually watch the news if I'm at home.

Example #2

I agree, but I also like to watch television. And I like to read. If I did not have a television I think I could get along. But I also think television can be educational. There are good television and there are bad ones. When I get tired of television and go into my room and read a good mystery or something. My family has a lot of books. I like cartoons too.....

1-- Example #1

I disagree because some people don't care about their body and I disagree about unplugging our television sets and if we weren't caring about our body we would not be here now because if there weren't no television children won't have an education and some people children run the street about a small amount of people talk to television people about the children education they would start playing education movie on your television.

APPENDIX E

Supplementary Scoring Materials

PAPER ANNOTATION

One way of keeping track of the content of papers is through annotation on either the margin of the score sheet or on a scrap page. You may not mark on student essays.

Marks for support and elaboration may appear as follows:

1.0	Example or Point
1.1	Detail for Example or Point
1.1.1	Elaboration for Detail for Example or Point

Letter code added to these numerical codes will describe the nature of the example or point.

S	for specific word (not necessarily empirical or real)
O	for opinion
S/O	for opinion but generally agreed to by people - folk wisdom
V	for vague
G	for generalization following support

Examples of Kinds of Support

WARRANTED OPINIONS(W)

There is too much violence on TV

Many shows are educational

Children might believe, imitate

Children would be outside getting exercise

Kids don't want to do anything but watch TV

Everyone will be bored/ have nothing to do, so instead they will

OPINIONS IN NEED OF FURTHER BACKING(O)

I like TV

There are many good shows

TV is harmful

Everyone will become alcoholics/criminals

TV is the only source of education

SPECIFIC SUPPORT(S)

Examples:
Sesame Street

My brother hit me

My sister learned her ABCs

I read that...

Recommendations:
Regulate TV
Do homework first
Watch TV as a family

BEST COPY AVAILABLE

APPENDIX F
SCORE SHEETS
PROGRAM EVALUATION AND ASSESSMENT
1983 IIEP WRITING RATER FORM

PACKET	RATER NUMBER	STUDENT ID NUMBER	SCHOOL CODE	GROUP/ QUARTER	IGE NUMB	SEQ
551	11	02435	437	05/02	3220	01

CIRCLE APPROPRIATE RATING

OVERALL RATING	FOCUS	ORGANIZATION
1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6

SUPPORT	MECHANICS/GRAMMAR
1 2 3 4 5 6	1 2 3 4 5 6

SENTENCE CONSTRUCTION	USAGE	PUNCTUATION/ CAPITALIZATION	SPELLING	PARAGRAPH FORMAT
0 1	0 1	0 1	0 1	0 1

ILLINOIS STATE BOARD OF EDUCATION

PROGRAM EVALUATION AND ASSESSMENT — SECOND RATING

1983 IIEP WRITING RATER FORM

HOUR MIN
STOP 1:57
START 1:30
TOTAL MINUTES 27

PACKET 182 RATER 14 GROUP 02 QUARTER 04

STUDENT	SCHOOL	SEQ	OALL	FOCUS	ORG	SUP	M/G	SEN	USAGE	P/C	SPELL	P/F
03294	680	01	5	5	5	4	5					
03124	478	02	3	4	3	3	3	1	1	1	0	1
02178	613	03	3	4	4	3	5					
00559	328	04	3	3	3	2	4					
00377	116	05	2	2	2	2	4					
00484	325	06	3	4	3	3	3	1	0	1	0	1
03693	276	07	1	1	1	1	2	1	1	0	0	1
01508	173	08	2	2	1	1	3	1	0	0	0	0
00131	105	09	2	2	2	1	3	1	0	0	0	1
00503	524	10	3	3	3	3	3	0	1	0	0	0

BEST COPY AVAILABLE

I disagree, because I think people can learn from television. There are many television shows nowdays that educate a lot of people, especially the younger ones.

I some way, I do agree though because besides the educational television shows there are are the ones that have too much violence. So, I think that instead of unplugging all of the TV sets in America, we should take off the shows that have too much violence, and put on the ones that are more educational.

BEST COPY AVAILABLE

BEST COPY AVAILABLE

APPENDIX G

FEEDBACK SHEET

Code Number Check Paper I

ELEMENT	RATING	COMMENTS
FOCUS	6	Last sentence summarizes the points made in the composition.
ORGANIZATION	4	The two points are in separate paragraphs. Uses "though" and "so" to signal the logic.
SUPPORT	2	Support is vague "many shows that educate a lot of people..." "too much violence".
GRAMMAR/MECHANICS	5	A sentence error probably due to copying the first draft incorrectly, a few minor errors.

Mechanical Skills below mastery

_____ Sentence Construction

_____ Usage

_____ Spelling

_____ Punctuation and Capitalization

_____ Paragraph Format

APPENDIX H

Logistics Schedule/Checklist

1. Sufficient copies of:

   30 training papers and feedback sheets -- 100 of each paper
   Computer generated scoring sheets -- 20,000
   15 pilot papers -- 100 of each paper
   24 check papers with feedback sheets -- 100 of each paper

2. Random assignment of 7,200 student papers and placement in 720 packets for scoring

3. Computer generated list of scorers indicating the packet numbers they are to score.

4. Miscellaneous -- overheads, chalkboards, setting up room, setting up distribution center
   Computer program for scoring

During Training

1. Distribution and collection of training and scoring materials
2. Score sheets to keypunch

Suggestions and Comments

1. An accurate scoring time per essay is necessary for proper scheduling.

2. The computer-generated lists with packet assignments will allow faster raters to go ahead and score additional packets without having to wait for slower ones.

3. The problem of waiting for packets could be alleviated by duplicating all student papers so that each scorer would have his/her own set. This may or may not be feasible depending on the number of papers to be scored.

BEST COPY AVAILABLE

APPENDIX I

EXAMPLE OF A "6" PAPER

Writing Assignment

It always strikes me as a terrible shame to see young people spending so much of their time staring at television. If we could unplug all the TV sets in America, our children would grow up to be healthier, better educated, and more independent human beings.

Do you agree or disagree? Explain and illustrate your answer from your own experience, your observations of others, or your reading.

Television does have a bad effect upon the children growing up today. The problem, however, is not one of content but of the addiction some children have to television. It can be a good way to relax and have fun if watched in moderation, but some children reach the level of watching television five or more hours a day. Clearly this is an extreme: it interferes with studying, outside play, and the exercise each child needs everyday. Watching television ½ hr to 1 hr a day can be beneficial in releasing tensions, but after a longer amount of time the child can not pry himself away from the TV set, and he loses sense of reality. Television tends to separate children from the outside world that they are preparing to live in on their own some day. Television is another world in the minds of children, and they are caught up in it.

Television reduces the creativity and imagination in children. They don't think for themselves - it does the thinking. They

BEST COPY AVAILABLE

no longer have the ability to imagine new and different worlds - it imagines for them. Children are reduced to the mentality of the programs to watch.; considering the low level of intelligence in programming today, it's no wonder that children are turning out less bright and imaginative. Watching t.v. all day, they have no time to create for themselves.

Television shuts off the communication between children and their families and between parents. There is no longer a great deal of conversation between child and parent, all eyes are glued to the set. T.V. is the automatic baby sitter - the parent doesn't have to play, talk or read with the child when he is occupied with a t.v. program. A lot of loving attention is lost on both sides. People lose a great deal of personal contact watching television - no one wants to be bothered. Children are growing up without loving parental attention more and more, and that is why they are turning out worse and worse.

If less television were watched in America, more bright happy children would result.

BEST COPY AVAILABLE

Glossary for Writing Assessment

analytical scoring - isolation and independent scoring of the characteristics or traits of good writing. Each paper is measured against specific external criteria.

check papers - (validation papers) - previously scored student essays which are rescored by raters to see if they remain consistent and accurate throughout the scoring process

criterion referenced scoring - the use of external standards or criteria to determine the level of writing ability as opposed to norms or standards based on the total pool of papers.

direct writing assessment - evaluation of writing ability based on an actual sample of writing

feedback sheet - paper given to raters which gives the check paper scores from the validation committee and the rationale for each

holistic scoring - overall impression of the effectiveness of a piece of writing. Papers are scored by relating the quality of the essay to other papers.

indirect writing assessment - evaluation of writing ability through the use of a multiple choice or objective test

interrater agreement - the similarity of scoring the student essays by two or more raters.

large scale assessment - the evaluation of thousands of student essays, usually at a state level

modes of discourse - type or method of writing which usually includes narrative/story telling, explanatory/reporting, and persuasive/argumentative

pilot papers - previously scored papers used to "test" whether raters are ready to begin scoring independently

primary trait scoring - focuses on specific rhetorical characteristic. It is based on the premise that all writing is done in terms of a specific audience. Papers are measured against specific external criterion.

prompt - writing assignment or topic

"rater drift" - term used to define the condition in which raters who are originally on target in using scoring criteria are no longer using it consistently

refereeing or third scoring - the rescoring of a student essay when the first two raters are far apart or when a crucial decision of promotion or graduation is based on the score.

BEST COPY AVAILABLE

table leaders - individuals who receive training prior to regular raters and are then used as assistants during the training session

training papers - previously scored (by the validation committee) student papers which are used to illustrate various score points for scorers or raters during a training session.

validation committee - group of experts who score student papers for the training sessions of potential scorers

writing across the curriculum - refers to instructional programs which employ the teaching of writing in various subject areas

Outline Guide for Writing Assessment

The following is a sequential outline of decisions and activities which need to be planned well in advance of a writing assessment.

1. Define what is and is not being tested

2. Decide on the best method for assessing the student writing

    a. Direct, indirect
    b. Holistic, primary trait, analytic

3. Determine the modes of discourse to be assessed

4. Choose/develop writing prompts

    a. Sufficient background for student to respond
    b. Unbiased
    c. Interesting
    d. Ageless-appeal to variety of maturity levels
    e. Topic information and directions for selection process
    f. Possible reminders for students to check their work
    g. Time restriction
    h. Directions on revision (checklist)
    i. Type of format (letter, essay, story, etc.)

5. Determine the rating criteria

    a. Consistency of criteria application
    b. Measure use for individual growth
    c. Breadth of scale (number of points above and below competency level)
    d. Precision of criteria which can be applied to one prompt or across a number of prompts
    e. Score format - single or combination of scores
    f. Use across or within grade levels
    h. Provision of information to students and teachers about writing strengths and weaknesses.
    i. Reflect policy decisions about the most important aspects of writing.
    j. Reflect results of research in defining writing components

6. Pilot and field test prompts to ascertain whether they will elicit the types of student responses necessary for the assessment.

7. Appointment of a validation committee to prescore student papers for training, pilot and check papers.

8. Recruitment and selection of raters

    a. State assessment may define specific criteria (writing samples, questionnaires, grade level, experience, etc.)

    b. Local assessment may have to utilize existing personnel.

BEST COPY AVAILABLE

9. Training Session (length of training determined by number of prompts)

   a. Introduction of raters to scoring guide and use of prescored training papers which represent various score points

   b. Rater practice

   c. Pilot papers given to raters to see if they are ready to score independently i.e., whether they can apply the criteria accurately as determined by validation commitee

   d. Check papers distributed to raters throughout scoring to determine consistent scoring

   e. Remediation for those raters not scoring to criteria

10. Check the technical quality of the assessment for interrater and validation paper reliability

11. Reporting student outcomes based on scoring criteria.

Illinois
State Board of
Education

100 North First Street
Springfield, Illinois 62777

Walter W. Naumer, Jr., Chairman
Illinois State Board of Education

Ted Sanders
State Superintendent of Education

EDUCATION IS EVERYONE'S FUTURE

Bulk Rate
U. S. POSTAGE
PAID
Permit No. 805
Springfield, IL

An Equal Opportunity/Affirmative Action Employer
Printed by the Authority of the State of Illinois
December 1984 12M 5-433B41 No.302